Foundations of Statistical Seismology

David Vere-Jones
Victoria University, Wellington
and Statistical Research Associates
New Zealand

1. Stochastic and Physical Models

1.1. Introduction

Nearly three decades ago, in 1979, I was asked to write an account of statistical modelling of earthquake occurrence in time and space.

Those three decades have seen great changes. Earthquake prediction has flowered, withered, and begun to flower again, albeit with a more cautious tone. We now have access to data of a scale and quality that would have been hard to anticipate 30 years ago, just as we have computing devices of a power and speed that would have been equally hard to anticipate.

In particular, the explosion of extensive, high-quality seismic data is a major reason behind the current increased interest in Statistical Seismology.

At just such a stage, it seems important to ask, as I was asked then, what is the purpose of stochastic modelling, in what has been traditionally viewed as an observational science, and how effective is it?

If Statistical Seismology is taken to mean the application of stochastic modelling ideas to Seismology, then this question is just a challenge to clarify the principles and purposes of Statistical Seismology itself.

1.2. What is a stochastic model?

The fundamental difference between a physical and a stochastic model, is that while the physical model seeks to understand and predict the process fully, the stochastic model accepts that some aspects of the physical process are out of range, at least for practical purposes, and must be replaced in the model by some unknowable and hence random process.

The main reason for making the uncertainties explicit, for building them into the model, is that it is only in this way that we shall be able to quantify the variability in the predicted outcomes.

The resulting stochastic model should reproduce those aspects of the physical phenomenon which are relevant and accessible to measurement, but may relegate the rest to dice-tossing or one of its contemporary avatars such as Brownian motion or the Poisson process.

1.3. Stochastic does not mean non-physical

However, just because a stochastic model treats some aspects of the process as random, that does not mean it is devoid of physical content.

More than three decades before my 1979 paper, Sir Harold Jeffreys (1938), who was a pioneer in inferential statistics as well as in geophysics, argued that, to be worthy of its name, every physical theory should contain within itself the means not only of predicting the relevant quantities, but also of predicting their uncertainties.

In our terminology, he was arguing that every physical theory should be based on a stochastic model.

But in adding to the theory the requirement that it should be capable of predicting the uncertainties, you do not take away the physics. You just add to it a further and often discomforting dimension.

1.4. Where does geophysics lie?

In classical physics, the uncertainties in the model are traditionally attributed to nothing deeper than observational errors. In quantum physics the situation is totally reversed: the uncertainties reflect a fundamental characteristic of the universe.

Geophysics, at the present time, occupies an uncomfortable middle ground.

General patterns of behaviour may be predicted qualitatively from physical theories, but the theories do not extend to the prediction of local earthquakes.

Our uncertainties include observational errors, but are by no means restricted to them.

A more fundamental difficulty is that we have only indirect observations on the physical processes taking place locally within the earth's crust. The processes themselves are complex, and for the present time out of range of direct observation.

Stochastic models of earthquake occurrence must somehow marry the limited physical theory to the limited data that bears directly on questions such as the initiation of a rupture and its development to a large-scale earthquake.

Under such circumstances, the requirement of being able to quantify the uncertainties in the model predictions represents a major and formidable challenge.

I believe it is fundamentally for this reason that the stochastic models that have been produced often appear to reflect the physical picture in such a limited way.

The underlying question for the geophysicist, then, is, " how can the observations and the physical picture be extended to allow a better quantification of the variability?"

2. Different roles for stochastic models

2.1 Two broad roles

Across their diverse fields of application, two broad roles for stochastic models may be distinguished.

The first is epitomized by statistical mechanics. Here the stochastic model plays an integral role in understanding the physical processes themselves.

In the second, by far more common, type of application, the stochastic model is used as a basis for planning, prediction or decision-making.

In this case, whether or not it fully represents the physical processes may not be the crucial aspect.

On the other hand, in such applications it is usually vital to know, not just a forecast value, but also its reliability. It is also vital that the model can be fully fitted to the available data. There is little practical use in having an excellent model which relies on information that cannot be accessed from the available data.

In my earlier paper I distinguished three broad classes of models, splitting the second class above into two: descriptive models and engineering models. Although I no longer like the terminology, I would like to examine each class briefly as it pertains to Seismology.

2.2. Descriptive models in Seismology: the G-R law

The aim in a descriptive model is to provide a recipe for producing data with the same broad features as those of the actual data.

In general, the simpler the model that will produce this effect, the more likely it is to be helpful.

Within Seismology, the canonical example would have to be the Gutenberg-Richter frequency-magnitude law. From the outset its purpose was purely descriptive, but the description took a left-hand turn.

Gutenberg and Richter, following a common habit among the physicists, first described their data in terms of numbers rather than proportions.

Then they used logarithm tables to base 10.

Finally they fitted a least squares regression line to the resultant numbers, thus obtaining

$$log_{10}N(M) = a + b(M - M_0) + E_M,$$

or equivalently

$$N(M) = 10^{a+b(M-M_0)+E_M}\,.$$

$N(M)$ here is the number of events in the data set which have magnitudes above M, M_0 is a magnitude threshold, and E_M is an error term which, to quote Harold Jeffreys once more, 'is quickly forgotten or altogether disregarded in physical theories.'

The tragedy to a statistician is that it is not a regression problem at all.

Just think how different elementary textbooks in seismology might appear if Gutenberg and Richter had phrased their discovery in terms of proportions rather than numbers, and in logarithms to base e rather than to base 10. Then they would have obtained

$$\log[\bar{F}(M)] = e^{-\beta(M-M_0)},$$

where $\bar{F}(M)$ is the proportion in the data set above magnitude M.

In this formulation, their discovery would have been clearly recognizable as a simple descriptive model for the distribution of magnitudes.

The pseudo-parameter 10^a disappears, being revealed as nothing more than a normalization constant (the total number of events above the threshold magnitude).

Moreover the term E_M is nothing like the error in a regression problem, but a quantity proportional to the discrepancy between the true and empirical distribution functions at the point M, a beast of a totally different character.

In my view, anyone pretending to the title of an up-to-date seismologist should be required on oath to forsake the use of the traditional form of the G-R law (other than in its historical context) and to persuade their colleagues to do likewise, to rid both text-books and current practice of a misleading anachronism.

Note that the model at this stage is purely descriptive. It is an empirical relationship. The reasons why the distribution should be exponential are nowhere related to a physical theory.

The second obvious example of a descriptive model is the Omori Law, at least when described, as suggested by Jeffreys (1938), as a Poisson process with time dependent rate of the form

$$\lambda(t) = A(c + t)^{-p},$$

where A, c and p are parameters and t is the elapsed time since the main shock.

This model is perfectly adequate for simulating a set of aftershocks with the same broad characteristics as a real set of aftershocks, and allows estimates to be made both of the parameters and of any predictions based on the model.

It may not fit an individual afershock sequence as well as the ETAS model, but in neither case is there is any explanation of why the power law form should be followed.

The ETAS model itself lies somewhere in-between classes. Primarily it is descriptive. Its components include:
- the G-R law (descriptive),
- the Omori law (descriptive)
- the exponential productivity law (descriptive),
- the spatial distribution of aftershocks (descriptive).

The only feature that (to me) has a conceptual rather than a descriptive basis is its *branching structure*: each event, whether background event or aftershock, producess offspring events according to the same formula.

2.3. Engineering (Utilitarian) Models

By this I mean models produced in order to answer a particular practical question in some planning, decision-making,or forecasting context.

There is broad overlap between such models and descriptive models The main difference is in the purpose of fitting the data. In a descriptive model the main purpose is simply to describe the data effectively. In an engineering model we want to put the model to some specified use.

Traditional uses of such models in seismology have been those relating to earthquake zoning, earthquake engineering design, and earthquake insurance.

But the major category now comprises models for probability earthquake forecasts.

The task is clear. It cannot be undertaken without a stochastic model. The question is whether the models are effective.

In formulating a stochastic model for any such practical purpose, some rough guiding principles can be helpful.

1. The level of detail of the model should match the purpose in view. There is no purpose in modelling detail that is not needed. Moreover a simple model is likely to be more helpful than a complex one in understanding and communicating the issues involved.

2. The model must be able to be estimated from the available data. No point in an excellent model that relies on unavailable data. This may mean restricting the

number of parameters. Commonly, 20 or 30 independent observations per parameter are needed to estimate each parameter even to moderate accuracy, although details vary hugely.

3. Even though following the physics may not be the main aim, a model which is based on a good, if simplified, physical picture, is likely to be safer for prediction than a model which is purely descriptive or ad hoc. A descriptive or ad hoc model just cannot be trusted outside the range of the data to which it has been fitted.

I see two broad situations in seismological studies where the models have this general character.

2.3(a) Analysis of data from an individual fault or a historical catalogue

Renewal, time-predictable, slip-predictable and stress-release models fall into this general picture. They have some physical plausibility, enough to satisfy (3) above, but their practical purpose is to provide estimates of the hazard on a given fault.

Point (2) is particularly relevant because the data is generally very meagre.

There is also a need to be careful with the model formulation to avoid internal inconsistencies. For example, one possible version of the time predictable model is

$$\log T_i = A + M_i + \epsilon_i \qquad (1)$$

where the $T_i = t_{i+1} - t_i$ are the times between events, the M_i are their magnitudes, and the ϵ_i are normally distributed errors.

The natural assumption of independent errors leads to a contradiction with the supposed boundedness of the stress level in time: without some negative correlations the fluctuations will increase beyond bound.

In the stress release model, instead of there being a fixed critical stress, the critical stress is treated as variable, having distribution function $\Phi(s)$ with density $\phi(s)$. The probability that the next earthquake occurs when the stress passes through $S, S+$ dS, but not before, is then given by

$$\Psi(S) = \phi(S)/[1 - \Phi(S)],$$

i.e. by the hazard function of Φ.

In applications, $\Psi(S)$ is commonly take to have an exponential form $\Psi(S) = Ae^{\lambda S}$, corresponding to the distribution function $\Phi(S) = 1 - e^{-A[e^{\lambda S} - 1]}$ which for $A << 1$ has a sharp peak at $(-\log A)/\lambda$. The stress-level is now Markovian, and the inconsistencies with the earlier model are avoided.

2.3(b). Models for background seismicity

The other group of models that play a somewhat similar role in a different context are the models for background seismicity such as the ETAS and Jackson-Kagan models.

The ETAS model has an important branching process interpretation, and is widely used as a basis for data-fitting, investigation of model properties (foreshocks, Båth's law) and as a diagnostic tool for revealing regions of anomalous seismic activity.

The Jackson-Kagan model was expressly designed for the purpose of providing a base-line model more realistic than the Poisson model but still simple.

The EEPAS model adds to the Jackson-Kagan model explicit prediction terms taken from logarithmic regression studies.

All three models can be defined by conditional intensities of deceptively similar form. For the full (space-time) ETAS model (Ogata, 1998),

$$\lambda_1(t,x,M) = f(M)\Big\{\mu(x) + A\sum_{i:t_i<t} \Phi(M_i - M_0)g(t-t_i)h(x-x_i)\Big\}.$$

For the Kagan-Jackson (1994) model,

$$\lambda_2(t,x,M) = f(M)H(t)\Big\{\delta + A_t\sum_{i:t_i<t} g(x-x_i)\Big\}.$$

For the EEPAS model, (Rhoades and Evison (1994)

$$\lambda_3(t,x,M) = \mu\lambda_0(t,x,M) + \sum_{t_i<t} f(M-M_i) h(t-t_i|M_i)g(x-x_i|M_i),$$

In these expressions f, g, h are all normalized to be probability densities, while $f(M)$ is the G-R law or one of its variants.

Here the similarities end.

Φ in the ETAS model is an exponential productivity term. It has to be balanced against the G-R term to determine the conditions for criticality. μ governs the background (independent) events and sets the overall spatial pattern. There are simple conditions for the existence of a stationary version, and when simulated from given initial conditions the model converges to its stationary form (ergodicity).

In the Kagan-Jackson model, the constant A_t is adjusted each time a new earthquake is added to the sum, to ensure that the total contribution from the bracketed term is unity and hence that $h(t)$ continues to denote the overall rate. When simulated, the model behaviour is heavily dependent on the initial condition, and the role of the ‘surprise events’ controlled by δ. It is not clear whether it can be linked to a stationary point process model, even if h is constant, and if so whether that model would be ergodic.

In the EEPAS model, λ_0 is first obtained from a model similar to the Jackson-Kagan model. The terms f, g, h in the sum are taken from logarithmic regression studies of the ratios of the seismic moment, time and space coordinates of an initial event to those of the events it anticipates. Again the model involves sequential renormalization, and it is not clear whether it can be associated with a stationary point process model.

Despite their varied backgrounds, all three models are successful in fulfilling what is required of them. However, they raise many further questions about the nature of the seismic regime and the models by which it can be represented.

2.4 Conceptual models

I mean here models that not merely describe but help to explain some physical phenomenon, as do the basic models in statistical mechanics.

Statistical models of somewhat this character have long played a role in the study of fracture mechanics, from the time of Griffiths (1926) and Weibull (1939) on.

Weibull, for example, attributed the variations in strength from otherwise similar laboratory specimens to the random distribution of microcrack lengths in the specimen. The Weibull distribution takes its name from his studies.

The branching process, percolation, and cellular automata interpretations of the earthquake process start from the underlying idea that, instead of progressing smoothly, as would a fault or fracture in a homogeneous elastic medium, the progress of an earthquake rupture is controlled by its essentially random progress from one weakness to another.

My own interest in this area revolved around the application of branching process ideas, leading to a stochastic model which predicted a G-R law with b-value around 2/3 in the critical case, and to tapered Pareto distributions ('Kagan distributions') when the process is subcritical.

It is remarkable that the same branching process concepts reappear in the ETAS model, lending credibility to one of Yan Kagan's old theses, that the distinction between the rupture itself, and the intervals between ruptures, are due more to the limitations of our perceptions and our recording instruments than they are to the physical processes.

It is also of interest to compare the roles of stochastic models for earthquake occurrence, such as the ETAS model or the branching model for fracture, with models for complex systems, whether stochastic, such as cellular automata, or deterministic, as in block-and-slider and many other mechanical models for fault systems.

Under a wide range of conditions, many show characteristic features of earthquake occurrence: a G-R law, long-range correlations, aftershock sequences and a form of Omori Law, etc. In this sense there may be no overriding reason for choosing one type of model over another.

Each provides a different type of insight into the circumstances under which these features can be produced. The merit of models such as the branching model for crack propagation, in my view, lies in the extent to which they can explain a complex phenomenon from simple premises.

I don't see much point in modelling a complex physical phenomenon by a model whose complexity approaches that of the original phenomenon, particularly when both may be adequately predicted by a simple statistical model.

3. On the statistical education of a geophysicist

In another early paper, I bemoaned the lack of time given to statistics courses in a typical degree programme in geophysics.

Traditional applied mathematics, physics, chemistry, geology - all these make up a really crowded programme for a geophysics student. And there is no time for statistics.

Until, that is, the student embarks on a project or thesis, when he or she is faced with the harsh realities of life in the form of a stack of observational data requiring interpretation, display, and the drawing of some kind of statistically legitimized conclusion.

There may be worse ways of learning statistics than being pitched in at the deep end. But it seemed a pity thirty years ago, and even more of a pity now if it is still true, that no serious attempt is made to incorporate statistics into the geophysics degree programme.

The advent of new and improved data, the growing interest in probabilistic forecasting and time-dependent hazard estimation, the powerful computing facilities now available to handle simulation and optimization techniques, all these point to a need to reassess the priorities, and to open up some pathway to inculcating a more mature form of statistical thinking among geophysics and especially seismology graduates.

However I emphatically do not advocate compulsory attendance at a cookbook statistics course. Many such courses are an insult to a mathematically literate student, and many geophysics students are more than a little mathematically literate.

Rather, the aim should be to acquaint students with the basic style of statistical thinking - probability models, their link to data, checking properties by simulation. Some familiarity with basic distributions and classical statistical tests will ultimately be needed, but is relatively easily learned. Familiarity with the basic style of thinking is harder to teach and more important.

My suggestions for a half-year course at around third year level might be something like this:

1. Take advantage of the modern statistical software which includes excellent

techniques for displaying data in many different forms. The importance of effective data display should be lesson 1.

2. The concept of a statistical model is best taught through simulation, generating random numbers according to the model specifications, from independent random samples to samples showing simple forms of chaining or dependence.

3. The empirical laws of seismology, and comparisons between actual and simulated data, offer plenty of scope for instructive and even rewarding discussions of statistical inference including both estimation and model testing. The aim in the latter should not be unquestioned obedience to 5% t-tests, but some understanding of the universal problem of trying to determine when a signal stands out above the noise.

4. Some introduction to simple stochastic processes, especially branching processes, simple Markov chains, AR models in time series.

At the MSc level and beyond, when it comes to training students to move into a field such as statistical seismology, there is a question as to whether one is looking to convert statistical graduates into seismology, or geophysics graduates into statisticians.

As a general rule, it is easier to do the mathematics (here I mean the statistics) first, and the more applied subject later, but both routes are possible.

I have found it easier to interest statistics students in seismology than seismology students in statistics. But the statistics students do not proceed far with seismology because they are lured away into careers in finance and the like.

The seismology students, on the other hand, fail to see in statistics a subject that warrants their attention when they could alternatively be going on enjoyable field trips in their own subject.

In any case, some attempt should be made to capture the interest of suitable students while they are still young enough to be impressed by challenges and ideas. Statistical seismology is surely an area where there is still important and exciting research to be done.

5. Conclusions and future perspectives

In this lecture I have enjoyed theopportunity to indulge my own prejudices and opinions.

I have tried to make, yet again, the case that stochastic models should not be seen as alternatives to physical models, but as extended versions of the physical models in which an attempt is made to explain the variability, or uncertainties, in the observations, as well as their basic causes.

At the same time stochastic models come in a number of guises and serve a number of purposes. These are likely to be more closely linked to physical considerations in some cases than in others.

— Descriptive models are no more or less than what they claim to be: a simplified description of the data.

— The majority of models, my so-called 'engineering models', are there to answer practical needs and should be judged in the first instance on whether they succeed in their stated tasks.

— Finally there are the more conceptual models, in which the aim is not only to describe but also to explain and understand the processes underlying some physical phenomenon.

The role of statistical modelling ideas in seismology has increased to the stage where more serrious attention should be given to the possibility of incorporating some serious statistical courses in the undergraduate and postgraduate statistical programmes. Better later than earlier, I think, and with the emphasis on statistical modelling, not on cook-book recipes.

In the meantime there is no shortage of new and important questions looming in statistical seismology. Let me just mention a couple of my own interests by way of somewhat far-out examples.

A few years ago I found a rigorously self-similar modification of the ETAS model, and suspect there may be a similar version of the EEPAS model.

There may be some way of linking these self-similarity ideas with the discovery, a few years ago by Brémaud and Massoulié (2001), of versions of the Hawkes (ETAS-type) processes, which run in the critical regime but *without immigrants*. It is possible that the Jackson-Kagan type models are linked to these.

I now believe that there may be some mathematical paradigm of the earthquake process which is exactly self-similar, and self-perpetuating. Five years ago I would have thought this ridiculous, but now I feel that nature may have beaten us to it in suggesting a remarkable new mathematical model.

Finally, the ‘rate and state’ friction ideas of Jim Dieterich seem to me to invite incorporation into a rigorous stochastic model, but the best way of setting up such a model is not yet clear, at least to me.

References

Vere-Jones, D. (1979). Distributions of earthquakes in space and time. In Chance and Nature, Ed. P.A.P. Moran, Australian Academy of Science, Australia ACT.

Jeffreys, H. (1939). Theory of Probability (1st Ed. 1939, 3rd Ed. 1961) , Cambridge, University Press.

Jeffreys, H. (1938). Aftershocks and periodicity in earthquakes. Beitr. Geophys. 53, 111-139.

Ogata, Y. (1998) Space-time point process models for earthquake occurrences. Annals of Institute of Statistical Mathematics, **50** 379-402.

Kagan Y. and Jackson, D. .D. (1994). Probabilistic forecasting of earthquakes. Geophys. J. Int. **143** 438-453

Rhoades, D.A. and Evison, F.F. (2004). Long-range earthquake forecasting with every event a precursor according to scale. Pure and Applied Geophysics, **161** 147-171

Brémaud and Massoulié (2001). Hawkes branching point processes without ancestors. Jl Applied Probability, **38** 122-135